



# The Lexicon-Grammar of French Verbs: a syntactic database

Christian Leclère

## ► To cite this version:

Christian Leclère. The Lexicon-Grammar of French Verbs: a syntactic database. The Lexicon-Grammar of French Verbs: a syntactic database, 2003, Tokyo, Japan. pp.33-46. hal-00192888

**HAL Id: hal-00192888**

**<https://hal.science/hal-00192888>**

Submitted on 29 Nov 2007

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

2003

## **The Lexicon-Grammar of French Verbs: a syntactic database**

Christian LECLÈRE (IGM, University of Marne-la-Vallée, France)<sup>1</sup>

### **Introduction**

The LADL (Laboratoire d'Automatique Documentaire et Linguistique)<sup>2</sup>, headed by Maurice GROSS from 1968 to 2002, aimed to classify all grammatical word classes in French according to their syntactic properties, and the distributional constraints that could characterize the sentences in which they occur. At the outset, it was essentially a linguistic approach, with no intention to build a tool for computational applications. But the way in which the description was formalized allowed us to incorporate the data within a general system capable of tagging very large corpora, analyzing texts and producing a syntactic description of sentences.

Our electronic dictionary provides information about the grammatical category (part-of-speech of each item), its possible inflected forms, and, in the case of verbs, a code indicating which syntactic class(es) it belongs to (COURTOIS 1997). For example, an entry like the following:

afficher V6 + 6, 35R, 38LD

indicates that the verb *afficher* has a V6 type of conjugation (i.e. together with the associated inflected forms), and that it belongs to syntactic classes 6, 35R and 38LD.

I shall briefly describe how the classification of verbs has been organized, and what kind of information it contains.

### **1. Syntactic description**

#### *1.1 General problem*

The Lexicon-Grammar is organized into a series of tables, each of them grouping items which share at least one main construction. This basic construction is

---

<sup>1</sup> I would like to thank Antoinette Renouf for her help. The translations provided are as close as possible to the French examples, rendering some of them rather unnatural.

<sup>2</sup> The LADL belongs to the CNRS (French National Research Center). It is now part of the Institut Garpard Monge at the University of Marne-la-Vallée (<http://infolingu.univ-mlv.fr>).

considered as the "defining property" of the item. For example, the verb *comparer* [compare] has the construction:<sup>3</sup>

$N_0 V N_1 \text{ avec } N_2$

where the preposition *avec* [with] can alternate with *et* [and]:

- (1) *John a comparé Jane (avec + et) sa mère*<sup>4</sup>  
John compared Jane (to + and) his mother

This construction has been considered characteristic of this verb, for a number of reasons, the main one being that it contains all the "arguments" that the meaning of the verb implies, the second one being that other verbs have the same characteristics -- like *marier* [marry], for example:

- (2) *Le prêtre a marié John (avec + et) Jane*  
The priest married John (to + and) Jane

This group of verbs constitutes a "natural class" of 129 "symmetrical" transitive verbs which are classified in the same "table" (Table 36S, see Figure 1 below).

### 1.2 Properties

Constructions (1) and (2) are obviously not the only ones for these verbs. For example, we can have a construction  $[N_1 \text{ et } N_2]_{N_0} \text{ se } V$ :

- (3) *John et Jane se marient*  
John and Jane get married (Lit. John and Jane marry each other)

where the two complements of (2) are in subject position (the verb is in pronominal form in this case). In each table of this type, various properties are encoded (in column) to indicate what other constructions are possible (Figure 1).

On the other hand, a verb like *permuter* [switch], which is in the same class because we can have sentences like:

*John a permuté la bouteille (avec + et) le verre*  
John switched the bottle (with + and) the glass

doesn't accept a pronominal construction of type (3),  $[N_1 \text{ et } N_2]_{N_0} \text{ se } V$ :

- (4) \**La bouteille et le verre se sont permutés*  
The bottle and the glass switched each other

---

<sup>3</sup>  $N_0$  is always the subject and  $N_1$ ,  $N_2$ , etc. the complements, prepositional or not.

N0				Table 36S	Pfx nég/source	N1								Prép		N2				N0 est Vpp (Prép N1)	N1 et N2 font un V-n	V-n entre N1 et N2					
N1 V avec N2	N1 se V avec N2	N1 et N2 se V	N2 V N1			N1 =; Nhum	N1 =; Npc	N0 lui V N1pc	N0 V N1 Loc N1pc	N0 V N0pc	N1 =; -Nhum	N1 =; le fait Qu P	N1 =; V-n	Ppv =; le	N0 V N1	Prép =; à	Prép =; de	N2 =; Nhum	N2 =; N-hum				N2 =; le fait Qu P	N2 =; V-n	Ppv =; lui	Ppv =; y	Ppv =; en
-	+	+	-			+	+	-	-	-	-	-	-	+	+	+	-	+	-				+	-	-	-	-
-	+	+	+	maquer	+	+	-	-	-	-	-	+	+	+	+	+	-	+	-	-	-	-	+				
-	+	+	+	marier	+	+	-	-	-	-	-	+	+	+	+	+	-	+	-	-	-	-	+				
-	+	+	-	marier	+	-	-	-	-	-	+	+	+	+	+	+	-	+	-	-	-	-	+				
-	+	+	-	mélanger	-	+	+	+	-	+	+	+	-	+	+	+	+	+	+	-	+	-	+				
-	+	+	-	mêler	+	+	+	-	-	+	+	-	+	-	+	-	+	+	+	-	+	-	+				
-	+	+	-	mésallier	-	+	+	-	-	-	-	+	+	+	+	+	-	+	-	-	-	-	+				
-	+	+	-	mixer	-	-	-	-	-	-	+	+	-	-	-	-	+	+	-	-	-	-	+				
-	+	+	-	opposer	-	+	+	-	-	-	+	+	-	+	-	+	+	+	+	-	+	-	+				
+	-	-	-	permuter	-	+	+	+	-	-	+	+	+	+	+	-	+	+	+	-	-	-	+				
-	+	+	-	rabibocher	-	+	+	-	-	-	-	+	-	-	-	-	+	-	-	-	-	-	+				

Figure 1: Extract of Table 36S

Instead, we would say (structure  $[N_1 \text{ et } N_2]_{N_0} V$ ):

- (5) *La bouteille et le verre ont permuté*  
The bottle and the glass switched (are switched)

which construction is not possible for *marier* [marry]:

- (8) \* *John et Jane ont marié*  
John and Jane married

Not all the verbs accepting sentences of type (5) are classified the same way. See for example:

- (6) *John et Jane flirtent*  
John and Jane flirt

The structure is the same as (5):

$N_0 V$  (With  $N_0 = Na \text{ et } Nb$ )

<sup>4</sup> "+", in parenthesis, means that there is a choice.

but there is no transitive construction as in (2) which (6) can relate to:

\* *Quelqu'un a flirté John (avec + et) Jane*  
 Somebody flirted John (with + and) Jane

So *flirter* cannot be in class 36S. On the other hand, (6) can be associated to (7):

(7) *John flirte avec Jane*  
 John flirts with Jane

Constructions (6) and (7) define Table 35 S (134 intransitive "symmetrical" verbs, see Figure 2).

N0 -: Num					Table 35S	Prép					N1												
N0 =: Nhum	N0 =: N-hum	N0 =: Nnr	N0 =: Vinf W	N0 =: V-n		N0 V	N0 est V-ant	N0 est Vpp	Prép =: dans	Prép =: à	Prép =: contre	Prép =: avec	Prép =: de + d'avec	N1 =: Nhum	N1 =: N-hum	N1 =: le fait Qu P	Ppv =: lui + y	Ppv =: en	Nhum V sur ce point	N0 est V-ant Prép N1	il y a V-n entre N0 et N1	N0 est en V-n avec N1	il V N0 W
+	-	-	-	-		+	-	-	-	-	-	+	-	+	-	-	-	-	-	-	-	-	+
+	-	-	-	-		+	-	-	-	-	+	+	-	+	+	-	-	-	-	-	-	+	+
+	-	-	-	-		+	-	-	-	-	+	+	-	+	+	-	-	-	-	-	+	-	+
+	-	-	-	+		-	-	-	-	-	+	+	-	+	-	-	-	+	-	+	+	-	+
+	-	-	-	-		-	-	-	-	-	+	-	-	+	+	-	-	-	-	+	+	-	+

Figure 2: Extract from Table 35S

### 1.3 "Defining" properties

As I said, the primary use of each given verb is defined by a main construction which defines all the verbs which have the same behaviour at the first level. The properties that are involved in the definition of tables are of three types: syntactic, distributional and semantic (LECLÈRE 2002).

As far as simple verbs are concerned, we distinguish 60 different classes of verbs (i.e. 60 tables) (M. GROSS 1975, J-P. BOONS, A. GUILLET & C. LECLERE 1976a, 1976b, A. GUILLET & C. LECLERE 1992). We have first taken into account the formal structure of the sentences in which each verb can occur. There are six of them:

$N_0 V$   
 $N_0 V N_1$   
 $N_0 V \textit{Prép} N_1$   
 $N_0 V N_1 \textit{Prép} N_2$   
 $N_0 V \textit{Prép} N_1 \textit{Prép} N_2$   
 $N_0 V N_1 \textit{Prép} N_2 \textit{Prép} N_3$

It's important to say here that, in all these analyses, the adverbial phrases are not taken into account, because they are not considered as characteristic of the verb. Everyone will agree that the adverbial phrases of place or time in:

*John a flirté avec Jane (dans le jardin + ce matin)*  
 John flirted with Jane (in the garden + this morning)

are not arguments that characterize the verb *flirt*. Nevertheless, at least some types of locative complements, considered as adverbial in traditional grammars, have been retained in the description of several classes of verbs. See:

*John a mis sa voiture dans le jardin*  
 John put his car in the garden  
*John a enlevé sa voiture du jardin*  
 John removed his car from the garden

Although they can take the same form as for other verbs, these locative complements do not have the same syntactic role when they are used with verbs like *mettre* [put] or *enlever* [remove]. There are dozens of verbs like these, for which these complements have to be considered as crucial arguments and not as adverbials.

Each of the  $N$  positions in the sentence structures above can be occupied by a noun or a sentence (noted  $Qu P$  [That S]). For example, the structure  $N_0 V N_1$  corresponds to three constructions:

- (1)  $N_0 V N_1$       Tables 32
- (2)  $Qu P V N_1$     Table 4
- (3)  $N_0 V Qu P$     Table 6

which can be illustrated respectively by the following three examples:

- (1) *John a abimé le livre* [John damaged the book]
- (2) *Que Jane vienne amuse John* [That Jane comes amuses John]
- (3) *John pense que Jane est folle* [John thinks that Jane is crazy]

The presence of a  $Qu P$  [that S] complement in the construction is one determining factor in the choice of the class which the verb belongs to and thus of the table in

which it appears. The verb *confier* [confide, entrust], for example, as in:

- (4) *Paul confie son problème à Marie*  
 Paul entrusts his problem to Mary  
 (5) *Paul confie à Marie qu'il doit partir*  
 Paul confides to Mary that he must go

will be classified as a sentence complement verb (Table 9) because of (5). Sentence (4) is considered as being derived from (5) (*That he must go is his problem*), and inventoried as such in Table 9. In contrast, the sentence:

- (6) *Luc confie sa valise à Max*  
 Luc entrusts his suitcase to Max

cannot be derived from a sentence complement, and so it appears in a table for constructions with nominal complements (Table 36DT in this case). It's interesting to note that, in many cases, the uses we distinguish between have different translations, but not always in the same constructions, as here for *confide* and *entrust*.

Such a purely formal classification, though useful at a first level, appears to be too coarse. To obtain more homogeneous classes of verbs, we need to associate the syntactic definitions with distributional properties; that is to say: specify what kind of preposition is possible, which features are attached to the different possible nouns in subject and complement positions, and so on. For example, *obéir* [obey] and *grouiller* [swarm] have the same construction  $N_0 V \text{Prép} N_1$ , but different prepositions, corresponding to two different tables:

$N_0 V \text{à} N_1$	Table 33	<i>John obéit à Jane</i> [John obeys Jane]
$N_0 V \text{de} N_1$	Table 35R	<i>John a changé de voiture</i> [John changed his car]

Other properties can be used to separate the uses of verbs more precisely, so that the final classes we obtain appear to be more or less homogeneous (when they are, we speak of "natural classes"). For example, the feature "obligatorily plural" attached to the direct complement of a few verbs (147 verbs) of structure  $N_0 V N_1$  leads us to put in the same table (32PL) verbs of which the meaning is roughly "gather things or people": *centraliser* [centralize], *collectionner* [collect], *rallier* [rally], *rassembler* [gather], etc.

Several properties can of course be combined to define a class. This is the case in Table 4, for example, where one can find a class of "psychological verbs" (*amuser* [amuse], *étonner* [surprise], *effrayer* [frighten], etc.

Structure:	$N_0 V N_1$
Properties:	$N_0 =: Qu P$
	$N_1 =: NhumObl (N \text{ "human" only})$

*Que John vienne (amuse + surprend + effraye) Jane*  
 That John comes (amuses + surprises + frightens) Jane

Note that the two properties do not have the same status: the direct object is obligatorily a human, but the subject can be a noun as well as a completeive (*This joke amuses Jane, John amuses Jane*).

One requirement in the selection of such properties is that they can be formally defined, from a linguistic point of view. Many of the features that we have chosen as properties are easily recognized because they are formally marked (like "obligatorily plural", which is generally marked by "s" or "x" in French). For others, it is necessary to use classification tests. For instance, a noun has the property *Nhum* (human) when it answers the question "qui ?" [who ?]: *Who is amused by John's coming ?*

In certain cases, only semantic properties can be used. The condition in this case (as in others, in fact), is that the intuition is "reproducible", whoever the native speaker is:

"Consensus among specialists is reached through experiments, but facts and experiments must be *reproductible*." (M. GROSS 2002:58)

For example, among those verbs with the construction  $N_0 V N_1$ , we have defined a sub-class, on the basis that the verb means "transformer en *V-n*" [transform into *V-n*]<sup>5</sup>. One can find, in this table, 131 verbs like *caraméliser* [caramelize], *gazéifier* [gasify] or *pronominaliser* [pronominalize]:

*John a caramélisé ce morceau de sucre*  
 John caramelized this piece of sugar = transformed it into caramel  
*On peut pronominaliser ce complément*  
 One can pronominalize this complement = transform it into a pronoun

#### 1.4 General processes of classification

To summarize, one can imagine a giant "super-table" which could take the form of Figure 3. The lines correspond to verbal entries (about 15,000 in French for simple verbs), and the different properties are in columns (about 300 of them have been tested). This super-table does not actually exist, but it represents what our work of classification has involved over several years. Theoretically, it represents 4,500,000 types of sentence. In fact, not all of them are studied for a given verb: to take a simple example, for an intransitive verb, it's clearly unnecessary to test all the properties of direct objects. Moreover, certain properties have been selected because of their relevance to a particular class of verbs but hold no significance

<sup>5</sup> *V-n* stands for any noun which is morphologically related to the verb (*caramel* / *caramelize*).



within other classes. Of course all the defining constructions have to be tested for each verb, before the table it belongs to can be decided.

	Definitional properties			Other properties			
	>>>>>>>>>Priority>>>>>>>>>						
	P1 (defining Table 1)	P2 (defining Table 2)	P3 (defining Table 3)	P4	P5	P6	P7
Verb 1	+	-	+				
Verb 2	-	-	+				
Verb 3	-	+	-				
Verb 4	+	+	-				

etc.

Figure 3: Theoretical general table

Let us consider only these defining properties<sup>6</sup>. They cannot be chosen so that they define separate classes of verbs. A verb generally has more than one defining property. To take a simple example: a verb often 'accepts' that one or the other of its complements are delete (*Paul mange un sandwich/ Paul mange* [Paul eats a sandwich/Paul eats]). We consider here that this is the same verb *manger* (there are many other verbs *manger*, with other meanings); so we do not create two entries, one in a table defined by  $N_0 V N_1$  and the other in a table defined by  $N_0 V$ . Instead, we give priority to the longer construction, because it is the one containing more information about the arguments of the verb. The fact that  $N_0 V$ , which can be a defining property for other verbs, is possible, will be regarded here as a simple property, and encoded in a column in the table defined by  $N_0 V N_1$  (Table 38L0 in this case)<sup>7</sup>. In the schematic case of Figure 3, property 1 (P1) has been given priority over P2, which has priority over P3. So the verb V1 will be classified within Table 1, in which property 3 will be noted in column, as P2 will be noted for V4, which is classified within Table 1 as well. On the other hand, V2, which has P3 but no other property, will be classified within Table 3. The consequence is that, if somebody is interested in a given property and wants to know the list of all the verbs which have it, s/he may have to look in different places. For example, the verbs which accept P3 are:

- all the verbs of Table 3 by definition (V2 here)
- all the verbs for which P3 is encoded "+" in other tables (V1 here, in Table 1).

<sup>6</sup> It should be noticed that I often use "property" and "construction" (or "sentence") interchangeably. The reason is that each property corresponds to a sentence. We consider that every feature has to be studied in context, the sentence being the minimal significant unit.

<sup>7</sup> This would not be the same for the verb *boire* [drink]: the sub-structure *John boit* [John drinks] having a special meaning ("John is an alcoholic"), it deserves an entry in a table defined by the structure  $N_0 V$ .

### 1.5 Splitting entries

While it appears that the verb *flirter* of (5) is the same as the one of (4), this is not always the case. A morphological verb has as many entries as it has uses that have been judged to be distinct. The distinction between two entries for the same verb, based on intuition at the beginning, has to be underpinned by appropriate properties. That becomes obvious when the different meanings correspond to different constructions. Take for example the verb *réaliser*: among its several meanings, it is easy to distinguish between one which can take a completive as direct object and another for which this is impossible:

*John a réalisé que Jane était partie* (Table 6)

John realized that Jane was gone (had gone)

*John a réalisé une œuvre d'art* (Table 32A)

John realized / created a masterpiece

But sometimes, two meanings (or more) can correspond to the same primary construction. In this case, we create two entries (or more) in the same table. Other properties encoded in this table allow us to justify the distinction. Look, for example, at the verb *communiquer* [communicate]. It has two entries in Table 35S (the same table of intransitive symmetrical verbs as *flirter* [flirt] above) (see Figure 4)):

*La chambre communique avec la cuisine / La chambre et la cuisine communiquent*

The bedroom communicates with the kitchen / The bedroom and the kitchen communicate

*John communique avec Jane / John et Jane communiquent (par e-mail)*

John communicates with Jane / John and Jane communicate (by e-mail)

N0 -: Num					Table 35S	Prép					N1												
N0 =: Nhum	N0 =: N-hum	N0 =: Nnr	N0 =: Vinf W	N0 =: V-n		N0 V	N0 est V-ant	N0 est Vpp	Prép =: dans	Prép =: à	Prép =: contre	Prép =: avec	Prép =: de + d'avec	N1 =: Nhum	N1 =: N-hum	N1 =: le fait Qu P	Ppv =: lui + y	Ppv =: en	Nhum V sur ce point	N0 est V-ant Prép N1	il y a V-n entre N0 et N1	N0 est en V-n avec N1	il V N0 W
+	+	-	-	-		+	+	-	-	-	+	-	-	+	-	-	-	-	+	+	+	+	+
-	-	-	-	-		-	-	-	-	-	-	+	+	-	+	-	-	-	-	-	-	-	-
-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 4 : Entries of *communiquer* [communicate] in Table 35S

Apart from the feature "human", attached to both subject and object in one case, and impermissible in the other one:

- \* *John communique avec la cuisine*  
John communicates with the kitchen
- \* *La chambre communique avec Jane*  
The bedroom communicates with Jane<sup>8</sup>

two other properties,  $N_0$  est *V-ant* and  $N_0$  est *V-ant Prép N<sub>1</sub>* confirm the difference:

- La chambre est communicante (avec la cuisine)*  
The bedroom is communicating with the kitchen
- \* *John est communicant (avec Jane)*  
John is communicating with Jane

At the present stage of classification, about 5,000 morphological French verbs yield about 15,000 different entries in 60 tables; that is to say, an average of 3 entries per verb; but of course a lot of them have only one entry, and some polysemic verbs may yield as many as 30 entries.

In conclusion, each entry for a verb in a table supposes that:

- the verb can be used in the defining construction of the table;
- it cannot be used with the same meaning in any more complex defining construction with higher priority;
- the construction in question is not a derived sentence. If this is the case, it is the source sentence that must be considered.

## 2. Support verbs and compound verbs

So far, I have considered only simple verbs. While carrying out our classification, we found that the description of the sentence predicate frequently required us to take into account not only the verb itself, but a combination of the verb and one or more nouns.

### 2.1 Support verbs

Let us consider the following examples:

- (1) *John projette de partir*  
John plans to leave
- (2) *John [a le projet] de partir*  
Lit. John [has the plan] to leave

---

<sup>8</sup> These sentences are possible if there is metonymy (*room / people in the room*). This is a question I cannot discuss here, but it is obvious that the processes consisting in asking systematic questions about such features as "human" and "non human" about every argument of the verbs is a fruitful way to investigate a lot of problems of this type and provides good examples to illustrate them.

It is clear that in (2), the predicative role is taken by the noun *projet* [plan] and not by the verb *avoir* [have]. It is the noun that decides the distribution of subjects and complements, in the same way as does the simple verb *projeter* [plan] in (1). The verb *avoir* [have] is only what we call a "verbe support" [support verb] (*Vsup*) of the predicative noun (*Npréd*). Such combinations [*Vsup Npréd*] are very numerous. Some of them correspond to a verb as in (1)-(2), but, in most cases, there is no equivalent simple verb (at least in French). See for example:

*John [fait un signe] à Jane*  
 John [makes a sign] to Jane  
*John [donne un rendez-vous] à Jane*  
 John [gives a rendez vous] to Jane (makes a rendez vous with)

Hundreds of such combinations have been itemized and have entries in special tables (see for example J. GIRY SCHNEIDER 1987, G. GROSS 1989 and R. VIVÈS 1983), organized in the same way as the tables of simple verbs (except that the entries are nouns).

## 2.2 Compound verbs and "frozen" sequences

An other case where it is necessary to consider compound predicates is where a verb is associated with one or more nouns, so that it is impossible to deduce the meaning of the expression from the meaning of the words of which it is composed. See the following sentences:

- (3) *John [brule les planches]*  
 John gives a spirited performance (Lit. John burns the boards)
- (4) *[Le rideau est tombé] sur cette affaire*  
 The curtain came down on this affair

The simple verbs *bruler* [burn] and *tomber* [fall] do exist as entries in tables of simple verbs (Table 32C and 35L respectively). These tables, of course, can only describe the proper meaning of (3) and (4). But we have here specialised uses of these verbs: in (3), nothing is really burnt, and in (4), there is no curtain. The sentences are not comprehensible if you only know the meaning of *bruler* [burn], *planches* [board], *rideau* [curtain] and *tomber* [fall]. The only way to describe such idiomatic cases is to take *V N<sub>1</sub>* [bruler les planche] and *N<sub>0</sub> V* [le rideau tombe] as complex units. We then create entries in tables of "frozen expressions". Other constraints can be observed in these complex units, in particular syntactic ones, as for example in:

*John garde / perd son sang-froid*  
 John keeps / loses his head (Lit. keeps / loses his cold blood)

where the determiner of *sang-froid* can only be a possessive (co-referential with the subject).

There are thousands of so called "frozen" combinations of this type, which do not obey the normal rules of simple verbs and deserve special treatment.

The electronic dictionary of LADL also contains other compound words, such as compound nouns like *perte de temps* [loss of time] or adverbs like *à toute vitesse* [in a hurry].

### 2.3 Example of the processes in the classification of a verb

We show here (Figure 5), with the verb *afficher*, an example of the way we have created entries for a verb and put them into appropriate tables, according to its syntactic and distributional properties. The sentences I give here are only examples of some of the sentences one can find in the tables (*C* = constraint noun, *Loc* = locative preposition and *V-n* = noun morphologically linked to the verb).

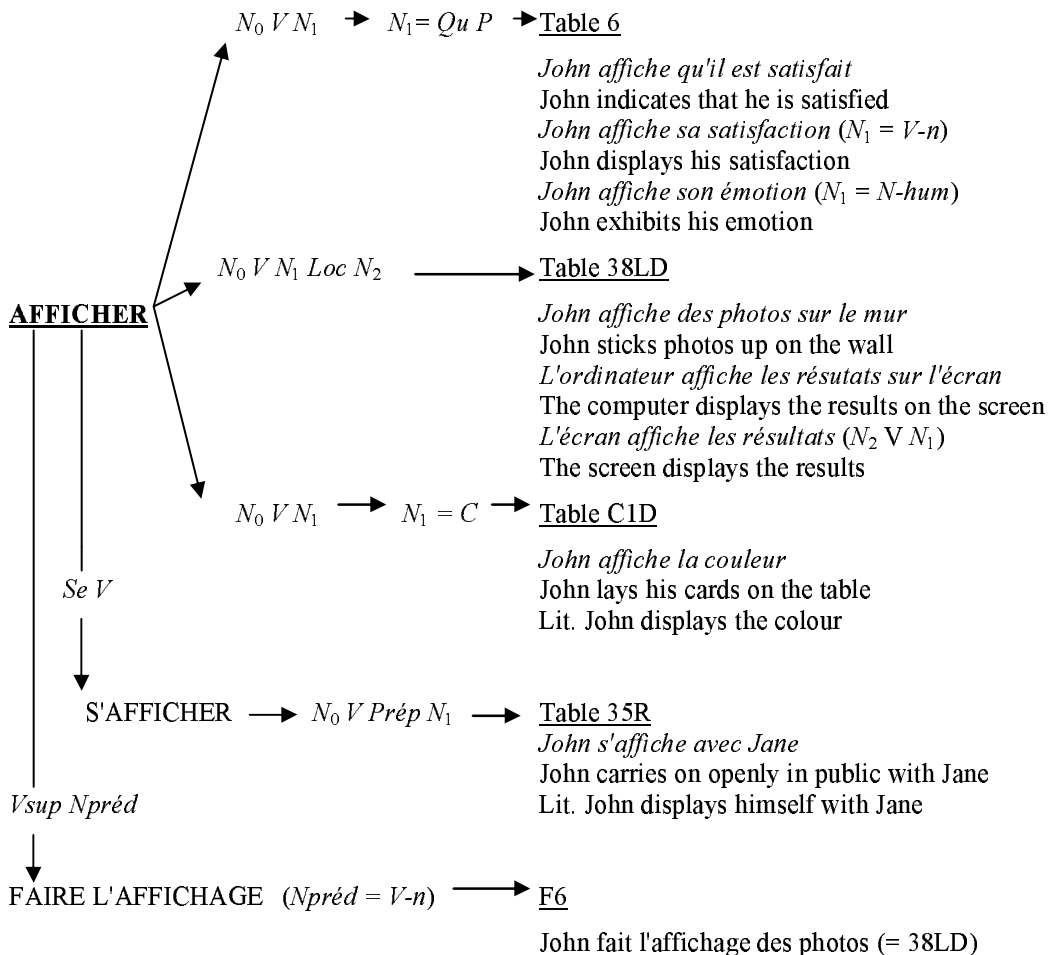


Figure 5: Part of the classification of the verb *afficher*

## 2.4 Results

Our **electronic dictionaries** make up what we call the 'DELA' system<sup>9</sup>. It contains:

- a dictionary of about 90,000 simple words (DELAS);
- a dictionary of corresponding phonetic forms (DELAP);
- a dictionary of more than 100,000 compound words (DELAC).

The inflected forms of simple words are automatically generated to produce the 'DELAf' dictionary.

In our **Lexicon-Grammar**, as far as verbs are concerned, we have entries in tables for:

- about 15,000 "free" constructions with simple verbs;
- about 25,000 "frozen" constructions with compound verbs;
- about 50,000 constructions with support verbs and predicative nouns.

As I said, with each verbal entry in our dictionary is associated the code(s) of the table(s) in which it is classified. This allows us to associate each verbal entry with all the main types of sentence in which it is likely to appear in texts.

## 3. Local grammars and graphs

The third part of our system consists of a series of "local grammars" which are formalized as FST (finite state automata). They have been created to describe sets of sentences which are used in a specific domain: expressions of dates, of temperature, stock exchange market reports (see T. Nakamura in this volume). I shall not describe these automata here. The interesting point here is that the Lexicon-Grammar, or at least part of it, can be converted into such graphs, and so applied to texts (see E. ROCHE 1999, S. PAUMIER 2001).

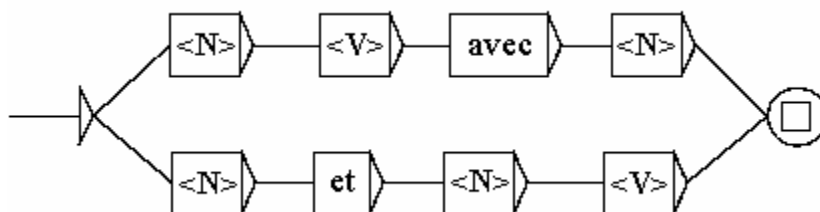
Schematically, a simple intransitive sentence can have the form:



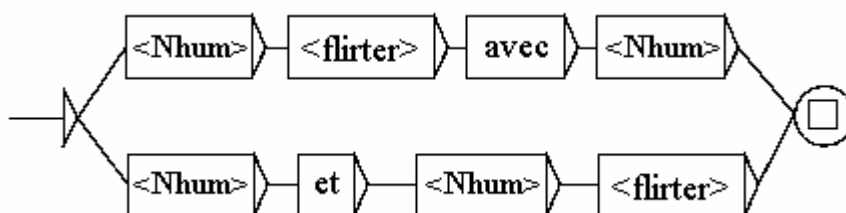
(<N>, <V> and <Prép> stand for any noun, verb or preposition respectively).

The defining property of Table 35S, for example, corresponds to a more precise graph:

<sup>9</sup> DELA stands for 'Dictionnaire Electronique du LADL' [Electronic Dictionary of LADL].



The properties encoded in Table 35S, corresponding to different constructions, can be converted into as many paths in the graph as there are "+" signs in the line of a given entry (the paths corresponding to "-" are of course eliminated). The verb *flirter* [flirt], for example, has the properties  $N_0 = Nhum$  and  $N_1 = Nhum$ . It can be associated with the following graph:



(<flirter> stands for all the inflected forms of the verb)

In theory, all the possible sentences described in the tables can be represented by graphs of this type. So, with each verb of the dictionary, or, more precisely, with each pair [V + code of table], we can associate a complex graph representing all the sentences we have retained as characteristic of the corresponding use of this verb. These graphs can be applied to a tagged corpus, but of course a lot of problems have not yet been solved:

- in practice, many properties (semantic, for example) cannot be exploited computationally;
- many derived constructions (imperative, for example) are not represented in tables;
- adverbial phrases, as well as various kinds of sequence which can be inserted at several places in sentences, are not taken into account (some of them have already been studied; see, for example, FAIRON 2000)
- etc.

## Conclusion

The systematic description of verbs (and other items) in syntactic tables is valuable, from a linguistic point of view, in raising many questions which have never been

examined. The final result constitutes a very large formalized database which is an invaluable set of information for researchers.

As for the computational applications, interesting results have already been obtained: dictionaries and various types of graph have already been incorporated into platforms like INTEX (M. SILBERZTEIN 1993, 1994) and UNITEX (S. PAUMIER 2002) for tagging and parsing very large corpora. The computational application of all the information contained in the lexicon-grammar raises some problems which are now being studied: it opens a lot of interesting avenues of research in the automatic treatment of texts, information retrieval, and even automatic translation (many other languages like English (M. SALKOFF 1995), Italian, Spanish, Korean are being described according to the same theoretical principles as for French).

## Bibliography

- BOONS, J.-P., A. GUILLET, C. LECLÈRE 1976a: *La structure de la phrase simple en français - Constructions intransitives*, Droz, Genève.
- BOONS, J.-P., A. GUILLET, C. LECLÈRE 1976b: *La structure de la phrase simple en français - Classes de constructions transitives*, Rapports de Recherche du LADL 6, Université Paris 7.
- COURTOIS, B. 1997: *Index du DELAS.v08 et du Lexique-Grammaire des verbes français*, Rapport Technique du LADL n° 54, tomes a et b, Paris, Université Paris 7.
- FAIRON, C. 2000: *Structures non-connexes. Grammaire des incises en français: description linguistique et outils informatiques*, Thèse, LADL, Université Paris 7.
- GIRY-SCHNEIDER, J. 1978: *Les nominalisations en français. L'opérateur "faire" dans le lexique*. Droz, Genève.
- GROSS, M. 1975. *Méthodes en syntaxe*, Hermann, Paris.
- GROSS, G. 1989: *Les constructions converses du français*, Droz, Genève.
- GROSS, M. 2002: "Consequences of the metalanguage being included in the language", in: B. E. Nevin (ed.), *The Legacy of Zellig Harris*, John Benjamins, Amsterdam/Philadelphia: 57-67.
- GUILLET, A., C. LECLÈRE 1992: *La structure de la phrase simple en français - Constructions transitives locatives*, Droz, Genève.
- LECLÈRE, Ch. 2002: "Organization of the Lexicon-Grammar of French verbs", *Linguisticæ Investigationes* XX:1, John Benjamins, Amsterdam/Philadelphia: 29-48.
- NAKAMURA, T. 2003: "Analysing texts in a specific domain with local grammars: The case of stock exchange market reports" (in this volume).
- PAUMIER, S. 2001: "Some remarks on the application of a lexicon-grammar", *Linguisticæ Investigationes* XXIV:2, John Benjamins, Amsterdam/Philadelphia: 245-256.



- PAUMIER, S. 2002: *Unitex - manuel d'utilisation*, Rapport de recherche, IGM, Université de Marne-la-Vallée, <http://infolingu.univ-mlv.fr>.
- ROCHE, E. 1999: "Finite state transducers: parsing free and frozen sentences". In *Extended finite state models of language*, A. Kornai (ed.), Studies in natural language processing, Cambridge University Press, Cambridge, UK: 108-120.
- SILBERZTEIN, M. 1993: *Dictionnaires électroniques et analyse automatique de textes: le système INTEX*, Masson, Paris.
- SILBERZTEIN, M. 1994: "INTEX: a corpus processing system". In *Proceedings of COLING 1994*: Kyoto.
- VIVÈS, R. 1983: *Avoir, prendre, perdre: constructions à verbe support et extensions aspectuelles*. Thèse de troisième cycle, LADL, Université Paris 7.
- SALKOFF, Morris. 1995: "On using the lexicon-grammar in a bilingual dictionary", *Lexiques-grammaires comparés et traitement automatique*, Presses de l'UQAM, Montréal: 311-325.

## INDEX

### Technical terms:

- Compound verb p.11
- Electronic dictionary p.1, 13
- Final state automata p.13
- Lexicon-grammar p.1, 13
- Support verb p.10
- Syntactic Tables p.1, 4, 7, 12

### Proper nouns:

- Gross, Maurice 1, 7

### *Author's address:*

Christian Leclère  
 32 rue Vitruve  
 75020 Paris  
 France  
[leclere@univ-mlv.fr](mailto:leclere@univ-mlv.fr)